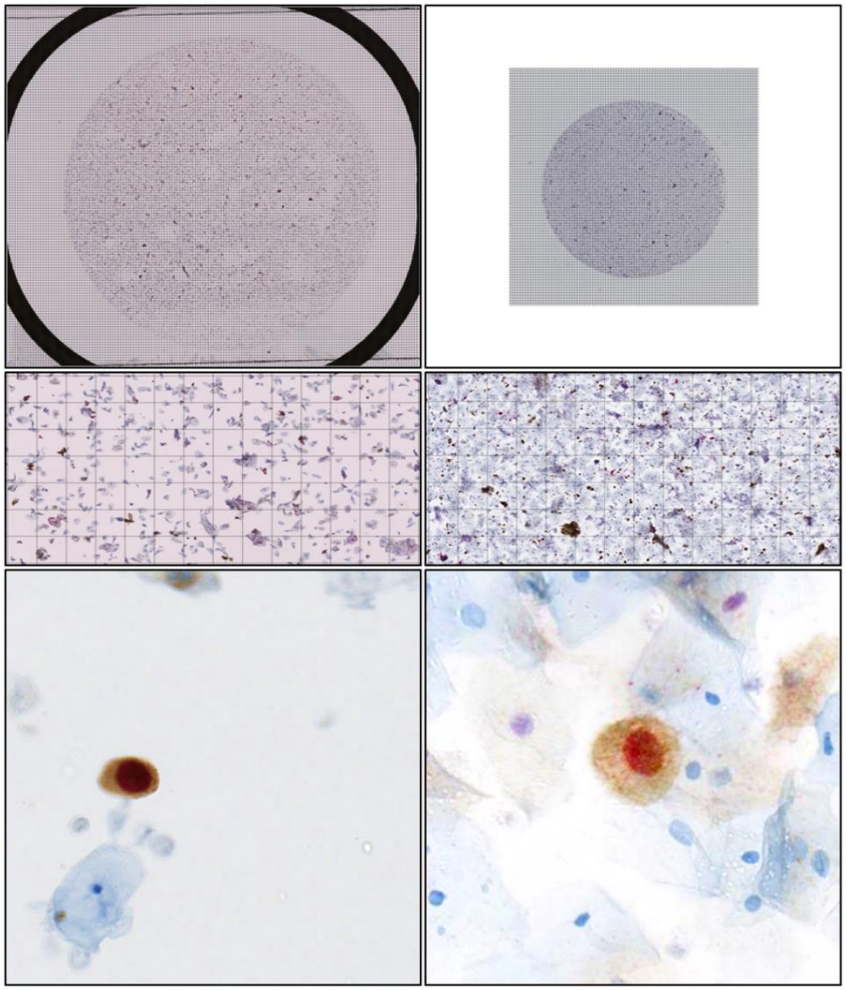
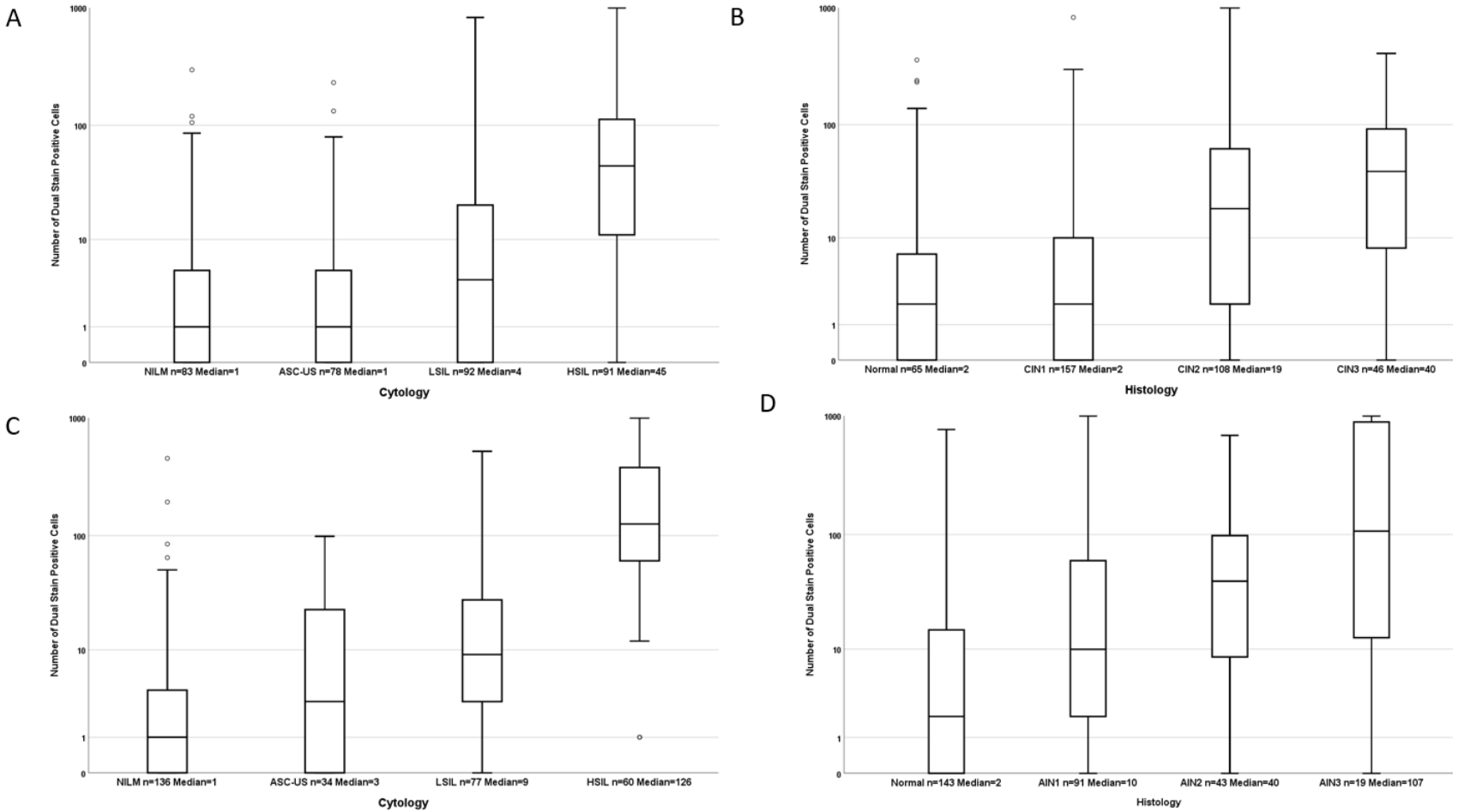
Supplementary Tables and Figure

Supplementary Figure 1: Dual stain cytology slide scanning



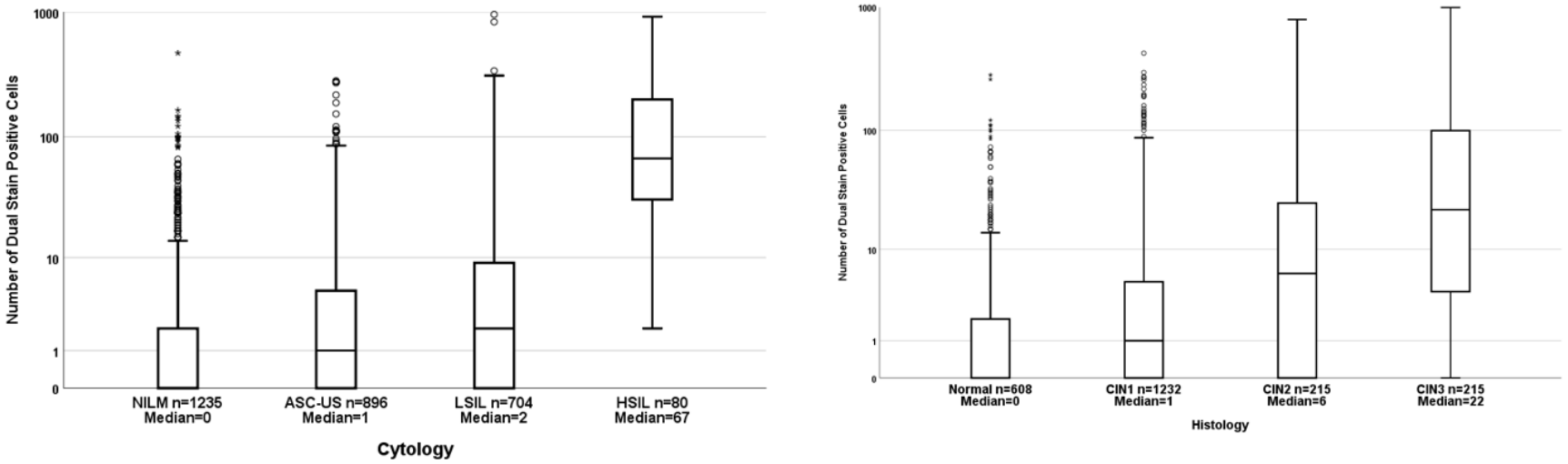
Whole slide scanning of ThinPrep (left) and SurePath (right) slides. The first row shows the slide at 0.42x magnification, the second row a more detailed view at 4x magnification, and the third row the tile view at 20x magnification.

Supplementary Figure 2: Number of dual stain positive cells by cytology and histology result in the Biopsy Study and the Anal Cancer Screening Study



A: Dual stain positive cells in cytology categories from the Biopsy Study; B: Dual stain positive cells in histology categories from the Biopsy Study; C: Dual stain positive cells in cytology categories from the Anal Cancer Screening Study; D: Dual stain positive cells in histology categories from the Anal Cancer Screening Study; Abbreviations: NILM, negative for intraepithelial lesions or malignancy; ASC-US, Atypical Squamous Cells of Undetermined Significance; LSIL, Low Grade Squamous Intraepithelial Lesions; HSIL, High Grade Squamous Intraepithelial Lesions; CIN1, cervical intraepithelial neoplasia grade 1; CIN2, grade 3; CIN3, grade 3; AIN1, anal intraepithelial neoplasia grade 1; AIN2, grade2; AIN3, grade 3

Supplementary Figure 3: Number of dual stain positive cells by cytology and histology result in in an HPV screening population in KPNC.



A: Dual stain positive cells in cytology categories Kaiser; B: Dual stain positive cells in histology categories from Kaiser
Abbreviations: NILM, negative for intraepithelial lesions or malignancy; ASC-US, Atypical Squamous Cells of Undetermined Significance; LSIL, Low Grade Squamous Intraepithelial Lesions; HSIL, High Grade Squamous Intraepithelial Lesions; CIN1, cervical intraepithelial neoplasia grade 1; CIN2, grade 3; CIN3, grade 3

Supplementary Table 1: Performance of automated dual stain in HPV16/18-negative women

Evaluation	Colposcopy referral	p-value (cytology/ manual DS)	Sensitivity	p-value (cytology/ manual DS)	Specificity	p-value (cytology/ manual DS)	PPV	p-value (cytology/ manual DS)	NPV	p-value (cytology/ manual DS)	Colposcopies per CIN3+ detected
Pap cytology (92 CIN3+)	1,453 (58.5%)	Ref	86.8% (80.3-93.2)	Ref	42.7% (40.7-44.7)	Ref	6.3% (5.1-7.6)	Ref	98.6% (97.9-99.3)	Ref	15.8
Manual DS (91 CIN3+)	1,148 (46.3%)	<0.0001/ Ref	85.8% (79.2-92.5)	0.8/Ref	55.5% (53.5-57.5)	<0.0001/ Ref	7.9% (6.4-9.5)	<0.0001/ Ref	98.9% (98.3-99.4)	0.6/ Ref	12.6
Automated DS (2 cells) (90 CIN3+)	922 (37.2%)	<0.0001/ <0.0001	84.9% (78.0-91.7)	0.7/0.8	65.0% (63.1-66.9)	<0.0001/ <0.0001	9.8% (7.8-11.7)	<0.0001/ <0.0001	99.0% (98.5-99.5)	0.4/0.7	10.2
Automated DS (1 cell) (94 CIN3+)	1,290 (52.0%)	<0.0001/ <0.0001	88.7% (82.6-94.7)	0.7/0.4	49.7% (47.7-51.7)	<0.0001/ <0.0001	7.3% (5.9-8.7)	0.005/0.04	99.0% (98.4-99.6)	0.4/0.7	13.7

Abbreviations: DS, dual stain; PPV, positive predictive value; NPV, negative predictive value; CIN3+, cervical intraepithelial neoplasia grade 3 or worse; Ref, reference

Supplementary Methods: Development of deep learning algorithms

For the general machine learning approach comprising the patient-, the tile- and the slide-abstraction level, we refer to Figure 1. To decompose slides into tiles, the OpenSlide library [1] is used with a tiles size of 384x384 pixels at 20x magnification, resulting in a physical geometry of 176.6um x176.6um (0.46um per pixel). As ThinPrep slides have a considerably larger sample area, they contain about 22,100 tiles, while SurePath slides contain about 8,500 tiles. For each tile, the CNN calculates the likelihood of the presence of at least one DS-positive cell. Among all DS-positive tiles, the median number of DS-positive cells was 1. Therefore, for simplicity, we refer to the number of DS-positive cells in the manuscript instead of tiles. Based on its comprehensively evaluated tiles, a patient is classified as positive if its total number of DS-positive cells on its slide exceeds a predefined number ("cut-point").

We developed the algorithms for ThinPrep and SurePath in sequence. We started the development in our two ThinPrep-based studies. Initially, we used support vector machines (SVM) to classify dual stain slides. However, that approach did not match manual performance (data not reported). We originally designed CNN4 as a simple convolutional neural network in analogy to AlexNet [2]. This network surpassed manual performance as desired. But subsequently, we saw that the network did not match manual performance in SurePath slides. Therefore, we extended the training set with a large number of tiles from SurePath slides, and we used a new state-of-the art network, IncV3, that was previously not available when we developed CNN4 for ThinPrep slides. Analogous to ensemble learning, [3] where different networks are used together, our approach resulted in two different network structures, based on the sequential development. Ensemble learning allows for continuous adaption of complex classifiers to evolving tasks over time like

in our example. We published a GitHub repository and created a web page at https://github.com/stcmedhub/dual_stain_dl with a source code description of the models and the installation instructions.

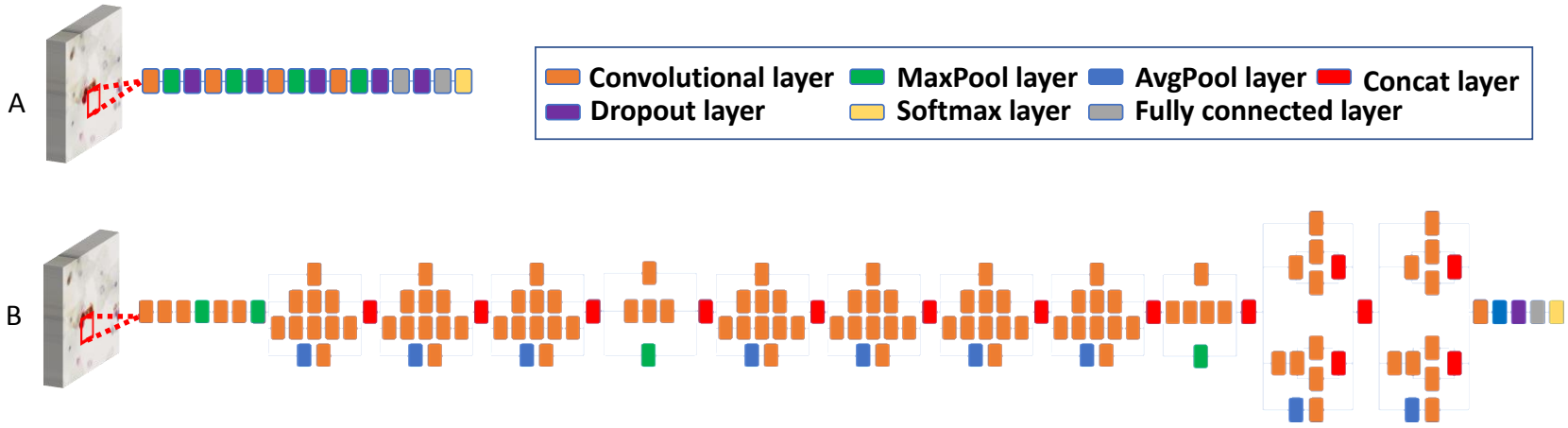
CNN4 algorithm

We developed deep learning algorithms for ThinPrep and SurePath slides, the two most-widely used liquid-based cytology technologies. We first developed a convolutional neural network for ThinPrep-based evaluating dual-stain cytology slides (CNN4). CNN4 has 4 convolutional layers [2] each with a subsequent pooling layer and two fully connected layers (Supplementary Figure 2). The final output layer was a softmax layer (>50% likelihood for a positive tile). Dropout layers were used to avoid overfitting during training. Rectified Linear Units (ReLUs) were used as activation function [4]. CNN4 was trained using Theano as a Python software development library supporting the development of deep-learning based algorithms [5]. Data preprocessing was applied to the training set by rescaling intensity values to a range of [0,1], applying mean subtraction and normalization by dividing each dimension by its standard deviation. We augmented the training set by flipping or rotating the images (leaving the underlying class unchanged). In total, we used 4,300 labelled images augmented to a total training set size of 68,800 images. The network was trained using backpropagation with Nesterov's Accelerated Gradient Descent [6]. For training, NVIDIA GTX 970 and GTX 980 GPUs were used.

Inception V3 algorithm

Since SurePath slides have a much higher density and somewhat different appearance of cells, a separate deep learning algorithm was developed for SurePath. We strongly enriched the training set with SurePath-based tiles and used InceptionV3 as a predefined

complex network structure consisting of 54 layers and Keras as a higher machine-learning software abstraction layer, allowing access to the TensorFlow backend. The "Inception V3" network (Supplementary Figure 2) was originally applied to the ILSVRC (ImageNet) classification challenge, to classify objects into one of 1000 categories [7]. Due to the binary classification problem of double-stain event detection, we changed the classification layers of the network to an output of two classes (>40% likelihood for a positive tile). We manually labelled a training set of 20,625 images, see Table DL3 for details. Data preprocessing was applied to the training set by rescaling intensity values to a range from 0 to 1, applying mean subtraction and normalization by dividing each dimension by its standard deviation. The networks were trained using the Stochastic Gradient Descent optimizer (SGD) with heavy data augmentation (rotation, color, shifting) but without transfer learning, resulting in a total training set size of 330,000 images.



Supplementary Figure 4. Comparison of the layouts of CNN4 (A) and IncV3(B). Data flow is from left to right, a 384x384 RGB image is used as input.

References

1. Goode A, Gilbert B, Harkes J, et al: OpenSlide: A Vendor-Neutral Software Foundation for Digital Pathology. J Pathol Inform, 4:27, 2013

2. Krizhevsky A, Sutskever I, Hinton GE: Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems 25:NIPS 2012, 2012.

3. Ju C, Bibaut A, van der Laan M: The Relative Performance of Ensemble Methods with Deep Convolutional Neural Networks for Image Classification. J Appl Stat 45:2800-18, 2018

4. Nair V, Hinton GE: Rectified linear units improve restricted Boltzmann machines. Proceedings of the 27th International Conference on Machine Learning (ICML):807–814, 2019

5. Bergstra J, Breulex O, Bastien F, et al: Theano: a cpu and gpu math expression compiler. Proceedings of the 9th Python in Science Conf (SciPy 2010), 2010

6. Sutskever I, Martens J, Dahl G, et al: On the importance of initialization and momentum in deep learning. International Conference on Machine Learning (ICML) 28:1139–1147, 2013

7. Russakovsky O, Deng J, Su H, et al: Imagenet large scale visual recognition challenge. International journal of computer vision 115: 211-252, 2015